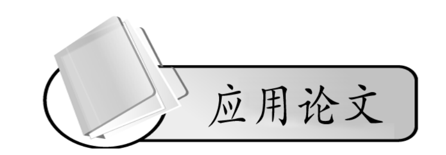


TeamDR:面向科研团队的数据知识库管理系统*

刘　峰[1,2,3]　黎建辉[1]　张　进[1]　韩　芳[1]　刘　昂[1]

[1](中国科学院计算机网络信息中心　北京　100190)

[2](中国科学院文献情报中心　北京　100190)

[3](中国科学院大学　北京　100049)

摘要:【目的】针对科研团队中分散的科研数据缺乏有效存储、管理, 无法复用的问题, 研发专门的数据知识库管理系统 TeamDR。【应用背景】TeamDR 是支撑课题组等科研团队用户完成科研数据组织、存储、管理及协作共享的便捷 Web 应用工具; 它采用 Java 为主要编程语言, 提供注册即可用的云服务版和本地安装版两个版本。【方法】针对科研多数据类型组织管理问题, 设计动态元数据内容模板, 同时为保证数据存储容量的可伸缩性、查询性能达到较高水平, 采用 MongoDB 作为存储设计。【结果】TeamDR 实现了科研团队数据存储与管理方面的重要功能: 如动态元数据模板、分级共享控制、元数据全文检索等, 试用反馈表明它满足了用户在数据存储管理方面的迫切需求。【结论】TeamDR 系统可以有效解决团队科研数据存储与管理、共享与协作、发现与关联方面的迫切基本需求。但在功能便捷性、完备性、扩展性方面存在进一步加强的空间。

关键词: 科研团队　数据管理　数据知识库　TeamDR

分类号: G250

1　引　言

数据密集型科研的蓬勃发展, 促使数据管理成为科研活动中密不可分的组成部分。进而推进了科学数据知识库(Data Repositories, DR)[1]的快速发展。

目前, 数据知识库主要分为机构数据知识库、学科数据知识库、多学科数据知识库以及特定项目数据知识库 4 类[2]。其中, 在数据服务的开放性方面, 学科数据知识库和多学科数据知识库由于面向广泛的科研群体, 开放性最强, 而机构数据知识库和项目数据知识库往往局限于相应机构或项目; 在服务学科领域的深度方面, 学科数据知识库面向特定学科领域, 且往往是长期服务, 表现出更强的系统化与专业化服务能力; 在服务学科领域的广度方面, 多学科数据知识库和机构数据知识库明显更有优势[1]。

然而就整体而言, 国内外数据知识库的建设重点仍集中在以存储加工数据和发表数据为主的公共服务 DR 方面。而针对相对分散的科研过程数据存储的数据知识库建设的重视程度还不够。

鉴于此, 笔者重点面向科研团队数据存储与管理问题, 采用动态模板设计和基于 MongoDB 的存储设计研发了 TeamDR[3]工具, 以期帮助科研团队有效完

通讯作者: 刘峰, ORICD: 0000-0002-5816-2067, E-mail: liufeng@cnic.cn。

*本文系国家"十二五"科技支撑计划项目"农村信息服务云存储与云计算技术研究及平台建设"(项目编号: 2013BAD15B02)、国家自然科学基金委重点基金项目"面向非常规突发事件应急管理的云服务体系和关键技术"(项目编号: 91224006)和中国科学院"十二五"信息化项目"科学数据资源整合与共享工程"(项目编号: XXH12504)的研究成果之一。

成数据管理工作。

2 需求与技术思路

2.1 国内外现状

随着国内外数字环境的发展，科学数据以及生产数据都迅速激增和积累，作为数据有效管理和研究的重要辅助手段，为满足数据人员的处理需求，各类数据管理和共享应用工具也随之出现[4]。

国外针对科研数据的存储管理不同需求也开发了一系列数据管理系统和存储工具，如非营利组织开放知识基金会开发的一个功能强大的开源数据门户平台软件系统 CKAN[5]、加州大学数字图书馆策管中心开发的新型的仓储服务系统 UC3 Merritt[6]、研究数据的存储和自由分享平台 Figshare[7]、存储生态学数据的权威数据仓储库 Dryad、DataStage、DataBank 以及 Scholar Sphere[8]等。

(1) Figshare 是一种新的分享开放科学数据的方式，其理念是可发现、可共享和可引用。在 Figshare 上，研究人员以可引述、可搜寻的方式发表数据，所有图片、媒体等上传，全都以 CC 标示，所有数据集则以 CC0 发布。Figshare 另一个重要特色是鼓励发布未发表的负面结果(Negative Data)和图片，通过这些数据，其他研究人员将不会重复工作[9]。通过采用 Amazon 基于云的数据管理系统，Figshare 保证了数据存储的安全性和可靠性。

(2) Scholar Sphere 是由宾州州立大学开发的开放仓储服务系统，师生和各类人员能够使用其收集和存储研究产出并且创建一个持久的、可读和可引用的记录，资源类型包括论文、演示文稿、出版物、数据集、学术杂志、数据、技术报告、音视频材料、年报、简报及其他成果等。研究者也能够利用此服务完成基金组织共享和管理研究数据的要求。

国内也涌现出一系列优秀的数据管理平台，如团队文档库、百度云盘、百会创造者、够快、简道云等。

(1) 团队文档库是由中国科技网出品的一个面向团队的文档数据协作和管理工具，对于科研团队、个人、项目团队、兴趣组等一系列多人组织，可以辅助有效进行文档型数据的协作管理工作[10]。文档库提供专属团队空间，用户可以加入多个团队，或构建多个团队，支持协作编辑、文档云存储等，方便共享，不怕丢失。

(2) 简道云是通过表单的方式实现数据搜集与管理的工具，用户可以通过其表单自定义的方式，搭建属于自己的数据管理软件[11]。主要提供数据搜集、数据管理、应用定制等功能，而且它的权限控制功能可以实现利用表单做内部管理。简道云既可以创建各种在线报名表单，反馈、调查表单，用来公开搜集数据，又可以基于企业内部的数据完成数据整理与协作，还提供丰富的图表组件，以供做好数据分析。

几种不同的数据管理应用的重点功能比较如表 1 所示:

表 1 数据管理应用比较

应用名称	Figshare	Scholar Sphere	团队文档库	简道云
数据存储方式	NoSQL	NoSQL	NoSQL	DB
大文件上传	1GB	2GB	1GB	×
元数据模板	9 种固定模板	多种固定模板	×	×
多粒度共享	×	×	√	数据集
数据组织	数据集	数据集	目录	√
数据发布	×	×	×	√
特色功能	全面计量统计、发布负面数据	多种开放性接口	文档协作	定制表单应用

比较结果表明，虽然现有的面向科研团队的数据管理工具在存储和功能方面各具特色，但在综合服务方面并未能很好地满足科研数据存储和组织管理的最基本需求。

2.2 应用需求

调研分析发现，当前科研团队在数据管理方面存在着迫切的应用需求，突出表现为以下方面:

(1) 科研数据分散在科研人员手中，无法持续存储积累，极易丢失。

(2) 科研过程与科研数据分离，缺乏有效的电子化整体存储管理工具。

(3) 科研过程管理不规范，科研数据的详细元数据缺失严重，数据无法有效理解、溯源、验证与重用。

(4) 科研数据共享及服务功能极不完善，数据共享模式单一，数据检索、同步困难，降低了科研效率。

2.3 定位与设计

鉴于以上分析与需求，TeamDR是定位于针对科研团队不同来源、多类型科研课题数据进行存储管理及

协作共享的工具，目前提供云端版和本地版两个版本。该工具重点专注于对科研数据的存储与管理；提供面向课题组的数据共享与协作；方便数据的有效发现与关联三个方面。

(1) 存储与管理

作为面向科研团队的数据知识库管理系统，TeamDR 专注于课题组各类科研数据资源的持续积累和有效管理，通过文件或关系数据库的存储实现数据资源长效保存和传承。构建课题组数据资源存储库的数据来源及存储类型结构如图 1 所示：

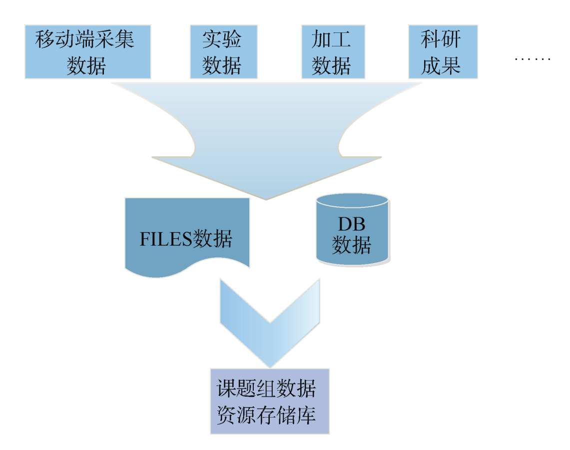


图 1　课题组数据资源存储库的数据来源及存储类型结构

在组织管理上，TeamDR 基于目录层级和数据集合进行结构组织，数据根据分类目录组织归档，再以面向科研主题的数据集合作为存储基本单位，对数据集元数据项及设置提供可动态定制功能，组织结构如图 2 所示。TeamDR 同时兼容关系型数据，支持数据在线协同编辑与检索，支持 Excel 表单与关系型数据库表间的导入和导出。

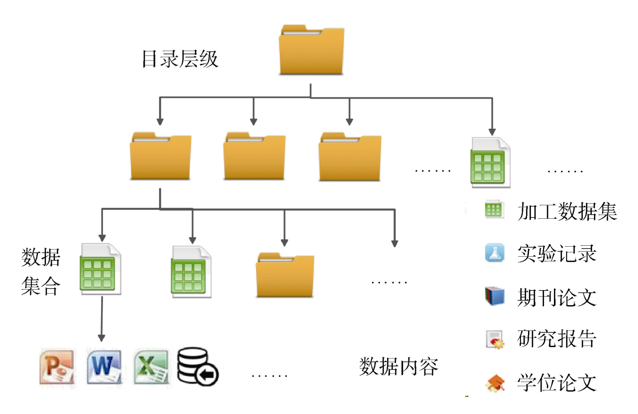


图 2　数据集合的目录层级结构

(2) 共享与协作

TeamDR 支持课题组的数据协作和共享，课题组成员可针对集合和关系型数据库协作创建数据内容，通过设定数据共享权限和相关标签标注，实现各种数据的灵活共享需求，并保证系统安全。

课题组数据资源库划分为个人空间、项目组空间以及课题组空间三部分，个人空间存储用户私有数据，项目组空间存储以项目(子课题)为单位的协同工作数据，课题组空间存储课题组公有数据。用户在各空间进行数据分享的权限设置如图 3 所示：

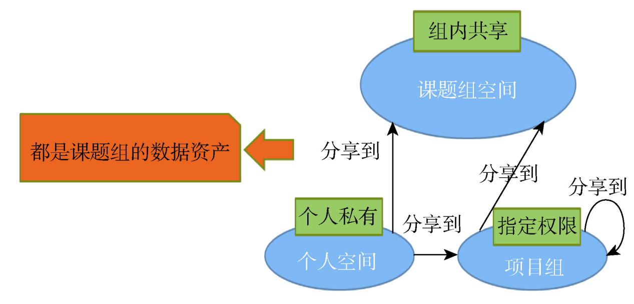


图 3　用户数据的权限共享方式

为满足课题组成员共享公开科研数据的需求，系统为科研用户提供数据网站的基本模板，课题组管理员用户可通过简单配置，即可以快速发布数据共享网站，如图 4 所示：



图 4　基于模板发布的数据共享网站

(3) 发现与关联

TeamDR 每类数据集合具有可定制的详细元数据信息；科研用户可根据特定的学科需求灵活定制不同类型的元数据内容模板，自由配置数据模板的各种录入属性，满足领域专属数据管理需求[12]。通过关联，可快速定位科研主题过程的原始数据，提供全文检索功能，让科研数据变得可理解、可验证、可追溯。元数据模板的定制流程如图 5 所示。

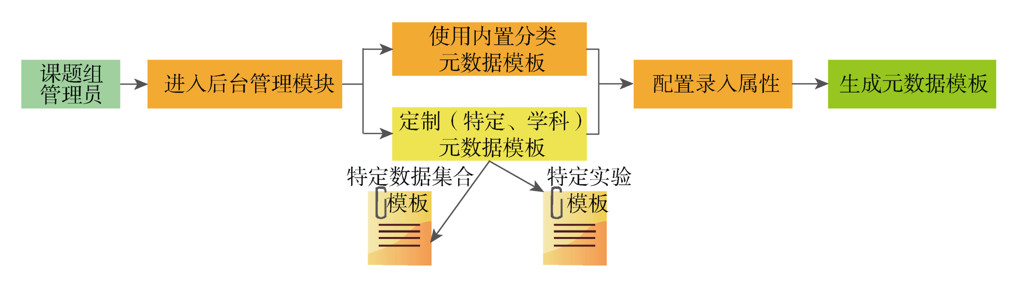


图 5　元数据模板的定制流程

2.4　技术路线

TeamDR 采用模块化的设计思路，将系统规划为多个服务模块，每个服务模块再细化为多个子服务模块。系统功能包括三个主要部分：数据资源的存储管理、数据资源的检索和数据资源的共享、辅助功能还包括数据发布和消息管理等。

系统整体设计采用 MVC 架构，对模型维护、数据展示、请求与响应进行分层处理。使用 Spring MVC 开发框架，采用功能强大并且容易集成的 Apache Shiro 安全框架进行权限控制。在用户交互上，借鉴主流 Web 在线资源管理应用的视图模式，提供类似于桌面资源管理器的操作界面，用户界面采用 Ajax 异步刷新技术，为用户提供数据资源浏览过程无跳转的体验；以 BootStrap 作为前端页面设计的基本框架，能够在不同分辨率的设备上实现良好展示，并保障前端界面安全友好。针对科研数据的多样性和不确定性，采用 MongoDB 存储元数据信息和结构化数据，并对文件型数据建立索引；使用云存储技术对文件型数据进行存储，便于维护和备份。

3　实现方案

3.1　系统架构设计

TeamDR 是一个 B/S 架构的应用系统，不仅具有易部署、客户端压力小等优点，还拥有典型的 3-Tier 分层结构，用户与数据的交互都通过中间服务层建立和进行，保证了低耦合的设计思想。图 6 是 TeamDR 的系统架构图，从逻辑上可将系统划分为三层：存储层、服务层和应用层。

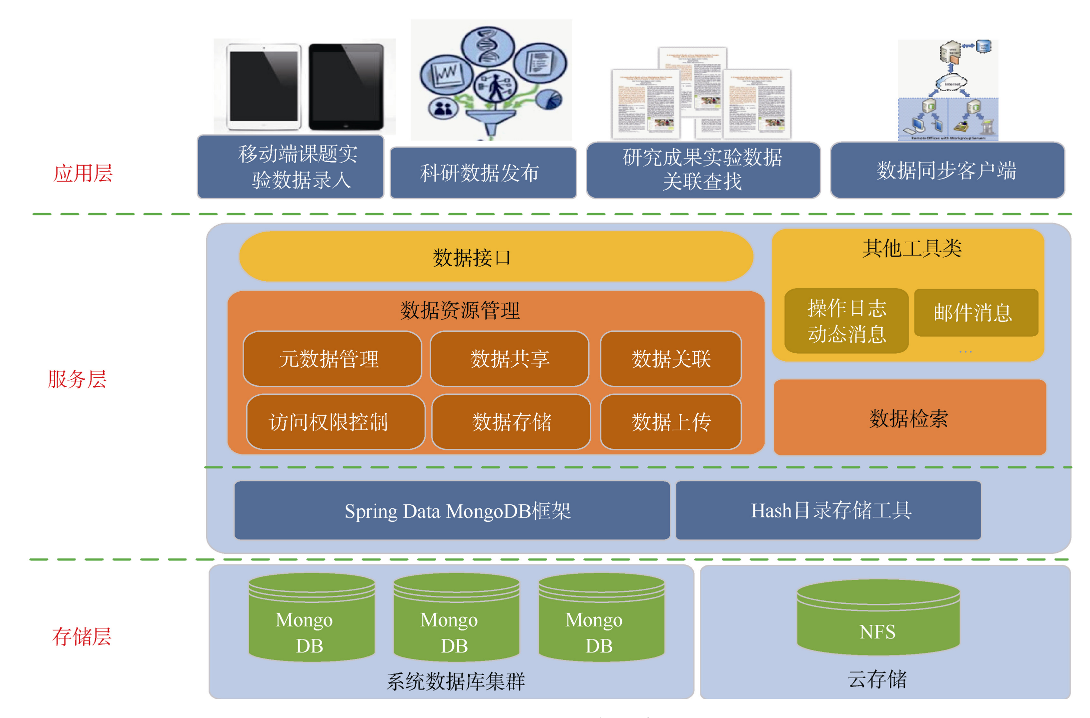


图 6　TeamDR 系统架构

(1) 存储层是系统数据存储管理的基础，用户的科研数据、元数据、以及系统数据等都在这层进行持久化。根据数据类型的不同，使用 MongoDB 数据库集群存储结构化数据，使用云存储文件系统管理非结构化文件型数据。这样可以充分利用各存储系统的优势，保证数据存储与获取的效率。

(2) 服务层是系统的核心业务逻辑层，围绕科研数据的存储、浏览和分享等主要功能分模块进行实现。其中数据访问使用 Spring Data MongoDB 构架支持服务层与 MongoDB 数据库的交互，而文件系统的存储采用三层 Hash 目录的存储操作接口。系统服务层的功能模块由数据资源管理、数据检索、数据接口、其他工具类等模块构成。其中，数据资源管理主要是对用户上传数据、浏览数据和进行数据分享等操作提供服务支持，同时保证各子模块间功能的协调。数据检索主要负责提供基于数据集合名称、数据类型和元数据的检索功能。数据接口主要向其他应用提供开放访问课题组数据资源的接口，包括获取目录下资源列表、获取数据集合元数据和获取数据资源等多种功能，以及受限数据集合的创建、数据的上传和发布等功能。其他工具类提供如记录操作日志、发送邮件等系统功能。

(3) 应用层是在基于以上服务层所提供的功能接口进行开发的应用程序，从数据采集、数据同步、数据查询到数据发布，旨在为用户提供更丰富的功能体验。目前，TeamDR 正在进行移动端数据录入应用，研究成果实验数据关联查找应用，数据同步客户端和科研数据发布对接等的开发。

3.2 关键技术设计

(1) 三层 Hash 目录存储设计

TeamDR 采用三层 Hash 目录存储方案，对每一个课题组建立一个根目录，课题组的数据文件存放在各自根目录下。由于不同硬盘格式能够存放的最大文件数目不同，同时考虑文件的索引和读取效率，不能简单将课题组内所有的文件都存放在一个目录下，因此建立一种基于 Hash 算法的目录索引方案，针对每个文件，将 MongoDB 生成的唯一索引作为文件的名称，同时将文件原属性作为元数据抽取出来存入 MongoDB 中；对文件名称使用三层 Hash 算法，计算出相应的三层目录路径，具体算法如下：

```
hash(filename)=HashCode(ObjectId)
path1=hash (filename)&255
path2=(hash(filename)>>8)&255
path3=(hash(filename)>>16)&255
path=Contact(path1,path2,path3)
```

①对 MongoDB 生成的唯一标示 ObjectId 做 HashCode 运算，得到文件名的 Hash 值；

②将文件名的 Hash 值与 255 做与运算，得到第一层文件夹名称；

③将文件名的 Hash 值向右位移 8 位，再与 255 做与运算，得到第二层文件夹名称；

④将文件名的 Hash 值向右位移 16 位，再与 255 做与运算，得到第三层文件夹名称；

⑤将得到的三层目录的地址连接起来，即可得到相对课题组存储目录的相对路径地址。

三层 Hash 存储设计可以避免单个文件夹内文件数目过多的问题，同时在文件的寻址上没有过多性能损耗。三层 Hash 存储设计是以课题组为单位，因此不会造成课题组之间的数据存储混乱无序，保证课题组的文件数据易维护、易备份。

(2) 基于 HTTP 协议的大文件上传设计

大文件上传问题一直困扰许多 B/S 架构的系统，通过 HTTP 协议单一连接上传，通常会出现连接超时等问题，即使对服务器端进行超时限制修改，如果在上传中出现错误，就会导致前功尽弃，需要重新上传。目前针对大文件上传的解决方案主要通过安装第三方组件实现，例如 Java Applet 方式、ActiveX 上传组件等，但效果并不理想。TeamDR 利用 HTML5 技术、MD5 算法[13]、文件拼接等技术，很好地解决了大文件上传问题。图 7 是 TeamDR 上传文件模块的流程图。

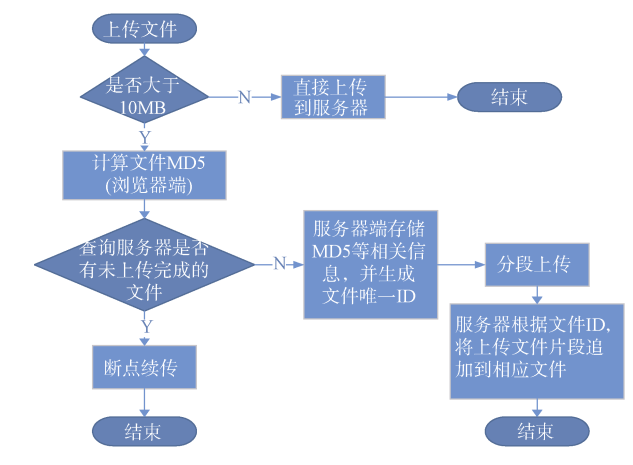


图 7 TeamDR 大文件上传处理流程

TeamDR 大文件上传处理的具体流程如下：

①用户在浏览器端选择上传文件后，通过 HTML5 File API 获取文件大小，如果文件小于 10MB，将直接上传文件；

②当文件大于 10MB 时，通过 HTML5 Blob API 计算文件的 MD5 值，获取文件的类型、大小。由于文件较大时计算文件 MD5 值花费时间较长，TeamDR 采用简易 MD5 值计算方法，截取文件前 64KB 数据和文件末尾 64KB 数据，计算这 128KB 数据的 MD5 值，具体计算方法如下：

S1 = file.slice(0,65536)

S2 = file.slice(file.size–65536,file.size)

MD5(file) = MD5(S1+S2)

此算法速度较快，并且可以与文件类型和文件大小联合唯一标识此文件。

③通过 MD5 值、文件类型、文件大小查询服务器上是否有未完成上传的文件。如果服务器端查找到此文件之前上传过，但是没有完全上传，则获取已上传文件的大小，并作为上传的起始字节点，继续上传文件。

④如果服务器没有找到相关文件，首先在服务器端存储文件名、文件大小、文件类型以及 MD5 值，并生成唯一 ID，便于上传失败后断点续传；浏览器获取唯一 ID 后，将文件分割成不超过 10MB 大小片段，并顺序上传。服务器收到文件片段后，根据唯一 ID 找到服务器上已上传的文件，并将新上传的片段追加到文件末尾。

(3) 基于 HTML5 的文档在线预览设计

文档在线预览功能可以实现用户无需安装相应软件，通过浏览器即可浏览相关文档，TeamDR 支持 Word、PowerPoint、Excel、PDF 等格式文档的在线预览。当前比较流行的文档在线预览方案大多基于 Flash，如豆丁网。

TeamDR 采用 HTML5 技术展示文档内容，不再受 Flash 限制，页面展示效果更加良好。并可让开发者利用 JavaScript 操作其对象，在网页上绘制图形图像[14]，而在 HTML 中则需要 Flash 插件实现。TeamDR 使用 Mozilla Lab 支持的 PDF.js 作为浏览器端呈现文档的工具，并在其上进行二次开发，与 TeamDR 的文件上传模块、数据存储等模块结合，提供流畅并且功能丰富的文档在线预览服务。具体流程如图 8 所示。

用户通过 TeamDR 系统上传文档后，数据库会存储文档的各种属性，如果文档是 Word、Excel、PowerPoint 等 Office 文档，TeamDR 会将文档转换为 PDF 格式，便于在线展示。为了提高用户体验，TeamDR 在后台设置了文档转换线程池，能够同时转换多个文档。

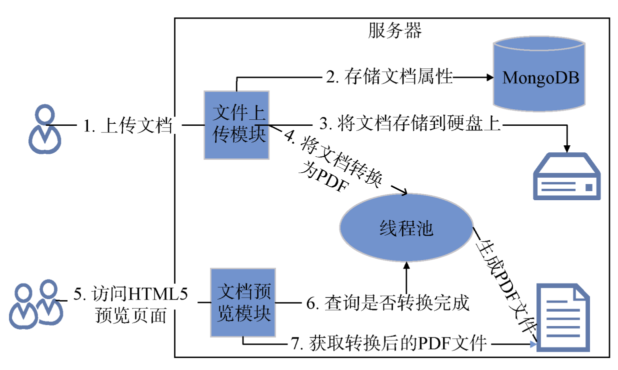


图 8　文档在线预览模块流程

TeamDR 目前支持 Office 文档和 PDF 文档的预览，并支持全屏展示等功能。页面采用 HTML5 技术，与各大主流浏览器兼容，并且摆脱了对 Flash 插件的依赖，在安全性设置较高的浏览器中依然能够正常展示。

4　应用效果

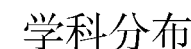
4.1　实现与测试环境

TeamDR 的云端版本的开发与测试主要依托中国科学院计算机网络信息中心的海云创新试验环境平台进行。此平台整体具备 1.2PB 的可扩展在线存储容量，单连接 WebService 传输性能达 150Mb/s 的网络。用户可进行自助式的服务申请，充分享有云计算服务环境的可伸缩、可扩展的特性，以及专业的 7×24 小时运维团队的技术支持。基于此平台的开发为 TeamDR 的部署和扩展都带来极大的便利，节约成本，为最终用户提供更稳定、更安全的服务。软件的开发主要基于 Java、JavaScript、HTML、CSS 等语言，选用主流稳定的开发框架如 Spring、BootStrap 等。

4.2　应用成效

经过对科研团队用户数据管理需求的明确，对关键技术进行研究与实现，TeamDR 云端版本已于 2015 年 8 月初正式上线运营，本地版本也在 8 月底提供下载和服务。产品上线后对特定科学数据库建库单位进行了初步宣传，并在科学数据大会、面向科研单位的交流研讨会上进行了推广，截至 2015 年 10 月底，云端注册用户 90 余名，本地版本下载 80 余次，云端课题组达 70 余个。目前的主要用户和课题组集中在化学、大气环境学科以及生物学科等领域，重点对中国科学院大连化学物理研究所分离分析化学重点实验室、生物技术部

1810 组，地理科学与资源研究所资源地理与国土资源研究室以及心理研究所核磁共振课题组等进行 TeamDR 使用培训，并部署了本地版正式使用。图 9 为 TeamDR 云端课题组的学科分布展示。



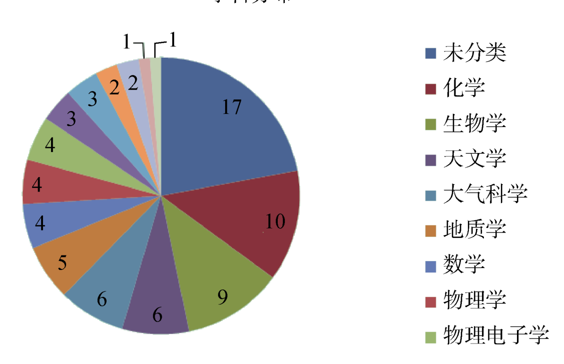


图 9 云端课题组的学科分布

另外，在中国科学数据领域最具影响力的创新大会——“2015 科学数据大会”上，首次参与宣传推广的“课题数据宝”产品即吸引了众多与会科学数据专家、学者和科研人员的目光，并最终获得大会专家委员会的一致认可，荣获了大会“最佳展示奖”。各个领域的科研学者也针对数据管理流程、安全性等方面为 TeamDR 提出许多建议和设想。

5 结论与展望

随着各学科领域的持续科学研究与发展，伴随科研过程而产生的试验数据、观测数据和文献数据等异质异构数据的规模在不断扩大，如何能够持续积累与有效管理这些科研数据资产，构建高质量的课题数据知识库成为越来越重要的问题。TeamDR 致力于帮助科研团队进行科研过程数据的有效存储与管理，将对解决这些问题带来极大的帮助。

TeamDR 在系统化功能完善方面还存在不足，需要进一步改进。主要表现在 TeamDR 与现有的权威数据知识库的接口对接方面；TeamDR 本地数据存储与云端数据进行实时同步方面；基于相机、GPS 定位、重力感应的 TeamDR 移动端数据采集与同步 APP 设计方面；TeamDR 数据资源版本控制方面等。

通过深入了解科研用户的需求，持续提升产品功能，TeamDR 会成为科研团队数据管理不可或缺的重要工具之一，同时笔者也希望 TeamDR 的研制能为科学数据知识库建设实践拓展一个新的思路和方向。

参考文献：

[1] 刘峰，张晓林，孔丽华. 科研数据知识库研究述评[J]. 现代图书情报技术，2014(2): 25-31. (Liu Feng, Zhang Xiaolin, Kong Lihua. Research Review on the Research Data Repositories [J]. New Technology of Library and Information Service, 2014(2): 25-31.)

[2] Pampel H, Vierkant P, Scholze F, et al. Making Research Data Repositories Visible: The Re3data.org Registry[J]. PLoS One, 2013,8(11). DOI: 10.1371/journal.pone.0078080.

[3] TeamDR [EB/OL].[2015-07-13]. http://www.teamdr.cn.

[4] 马建玲，曹月珍. 研究数据管理工具发展研究[J]. 图书馆学研究，2014 (15): 40-47. (Ma Jianling, Cao Yuezhen. Research on the Development of Research Data Management Tools[J]. Research on Library Science, 2014(15): 40-47.)

[5] CKAN [EB/OL]. [2015-11-08]. http://ckan.org/about/.

[6] UC3 Merritt [EB/OL]. [2015-11-08]. https://merritt.cdlib.org/.

[7] Figshare [EB/OL]. [2015-11-08]. http://figshare.com/.

[8] What is Scholar Sphere [EB/OL]. [2015-11-08]. https://scholarsphere.psu.edu/.

[9] 张静. Figshare 平台与 CNKI 学术图片库比较分析[J]. 科技与出版，2015(1): 63-66. (Zhang Jing. Comparative Analysis of Figshare Platform and CNKI Academic Picture Library [J]. Science-Technology & Publication, 2015(1): 63-66.)

[10] 科研在线团队文档库[EB/OL]. [2015-11-08]. http://ddl.escience.cn/.(Online Team Document Library of Scientific Research[EB/OL]. [2015-11-08]. http://ddl.escience.cn/.)

[11] 专业的数据收集管理工具：简道云[EB/OL]. [2015-11-08]. https://www.jiandaoyun.com/. (Professional Data Collection and Management Tools: JianDaoyun [EB/OL]. [2015-11-08]. https://www.jiandaoyun.com/.)

[12] 李瑞. 基于语义分析的元数据管理工具的设计与实现[D]. 武汉：华中科技大学，2012. (Li Rui. Design and Implementation of Metadata Management Tool Based on Semantic Analysis [D]. Wuhan: Huazhong University of Science and Technology, 2012.)

[13] 毛熠，陈娜. MD5 算法的研究与改进[J]. 计算机工程，2012, 38(24): 111-114. (Mao Yi, Chen Na. Research and Improvement of MD5 Algorithm [J]. Computer Engineering, 2012, 38(24): 111-114.)

[14] 夏翠娟, 张燕. 图书馆移动阅读服务的新契机: HTML5 和 CSS3[J]. 现代图书情报技术, 2012(5): 16-25. (Xia Cuijuan, Zhang Yan. The New Chance of Library Mobile Reading Services: HTML5 & CSS3 [J]. New Technology of Library and Information Service, 2012(5): 16-25.)

作者贡献声明:

刘峰: 提出整体思路与框架, 参与整体方案设计、内容分析, 论文统稿和修订;
黎建辉: 提出论文完善思路, 参与内容分析和论文修订;
张进: 系统实现框架设计与整理, 关键技术研究整理, 论文相关章节撰写和修订;
韩芳: 国内外应用现状调研, 系统定位与设计等, 论文相关章节撰写和修订;
刘昂: 系统关键实现技术研究, 论文相关章节撰写和修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据[1-2]见期刊网络版 http://www.infotech.ac.cn; 支撑数据[3]由作者自存储, E-mail: hanfang @cnic.cn。
[1] 刘峰, 黎建辉, 张进, 韩芳, 刘昂. CloudService_url.云端版服务相关链接.
[2] 刘峰, 黎建辉, 张进, 韩芳, 刘昂. Standalone_url.本地安装版相关链接.
[3] 刘峰, 黎建辉, 张进, 韩芳, 刘昂. DisciDistri.xls.云端版使用学科分布表.

收稿日期: 2015-09-29
收修改稿日期: 2015-11-16

TeamDR: A Data Repository Management System for Research Teams

Liu Feng[1,2,3] Li Jianhui[1] Zhang Jin[1] Han Fang[1] Liu Ang[1]
[1](Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)
[2](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)
[3](University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: [**Objective**] This study aims to effectively storage, manage and reuse scientific data with the help of a specialized data repository management system, the TeamDR, for the research teams. [**Context**] TeamDR is a Web tool helping scientific research team members organize, storage, manage and share data. It was developed by Java and offered cloud-based and standalone services. [**Methods**] We first designed a dynamic metadata template to organize and manage scientific research data. MongoDB was then adopted to improve data storage capacity and query performance. [**Results**] TeamDR stores and manages the scientific research data effectively with the support of, dynamic metadata template, categorized sharing control, and full-text search of metadata. Users' feedbacks show that TeamDR meets the demands of scientific data storage and management. [**Conclusions**] TeamDR effectively addresses the issues of scientific data storage and management, data sharing and collaboration, data discovery and linking. However, this system's usability, completeness and extensibility could be further improved.

Keywords: Scientific research teams Data management Data repositories TeamDR